

様式 7 (Form 7)

## 学 位 論 文 要 旨

Dissertation Abstract

学位請求論文題名 Dissertation Title

方向統計学におけるノンパラメトリック統計解析

(和訳または英訳) Japanese or English Translation

Nonparametric statistical analyses for directional statistics

人間社会環境学 専 攻 (Division)

氏 名 (Name) 鶴田 靖人

主任指導教員氏名 (Primary Supervisor) 寒河江 雅彦

(注) 学位論文要旨の表紙

Note: This is the cover page of the dissertation abstract.

## Abstract

This dissertation aims to discuss the theoretical aspects nonparametric statistical analyses such as the kernel density estimation (KDE) and the local polynomial regression (LPR) for directional data. Directional data are the observations that have a periodic variation such as energy demands for twenty-four hours. This dissertation provides the theoretical properties of the KDE and LPR employing a kernel class proposed Hall et al. (1987). Its KDE has some good properties: (i) The convergence rate of the mean integrated squared error (MISE) of this is improved by employing a  $p$ -th order kernel that can vanish the lower moments than  $p$ -th. (ii) The estimators of the optimal smoothing parameter of this are consistent with its parameter. Its LPR has some advantages: (i) Raising the order  $p$  of its polynomial improves the convergence rate of the conditional weighted MISE of this. (ii) Even if this has multivariate linear-torus explanatory variables, the convergence rate of its conditional weighted MISE can be provided.

## 学位論文要旨

### 序論 (1 章)

経済分野において、財の需要量 (供給量) の変動は、1 年周期・1 週間周期・24 時間周期などといった周期的な変動を持つ。このような周期的変動を持つデータを角度データと呼ぶ。また、気象学における風向や生態学における動物が移動する方向などの角度観測値も角度データの 1 種である。

実数値データを解析するために発展してきた従来の統計学は、データは実数直線  $\mathbb{R}$  上で定義された確率分布に従うという仮定の下で、検定や回帰など様々な統計手法を提案してきた。しかし、角度データは、実数値データのための統計手法をそのまま活用できないという問題を持つ。この問題は、角度データの位相と実数値データの位相が異なるために生じる。そのために、角度データを扱うための統計手法を研究する新しい統計学の分野が生まれた。それが方向統計学である。方向統計学は、角度データは単位円周上で定義される確率分布に従うという仮定の下で、検定や回帰といった様々な独自の統計手法を提案してきた。

本稿の研究目的は、方向統計学において未成熟な分野であったノンパラメトリック統計解析の理論を整備することである。本稿では、ノンパラメトリック統計解析とは、統計モデルにカーネル関数と呼ばれる重み関数を用いた統計手法のことを指す。

ノンパラメトリック統計解析の特徴は柔軟な統計モデルを構築可能なことである。その特徴からノンパラメトリック統計解析は複雑な構造を持つデータに関してもその構造を推定可能である

という長所を持つ。周期的変動を持つ経済データの多くは社会の複雑なメカニズムによって発生しているので、方向統計学におけるノンパラメトリック統計解析の理論研究の発展は、このような経済構造を解明可能にする。

本稿では、ノンパラメトリック統計解析のうち、カーネル密度推定量 (KDE: kernel density estimation) とノンパラメトリック回帰の理論的性質の理論研究を行った。本稿の主要な貢献は、Hall et al. (1987) が提案した Hall 型カーネル族をこれらの手法に適用することで、方向統計学における KDE・ノンパラメトリック回帰も、実数値データを解析するための通常の KDE・ノンパラメトリック回帰と同様な理論的に良い性質を持つことを明らかにしたことである。

次に、本稿の流れを説明する。1 章では従来の通常の KDE・ノンパラメトリック回帰・方向統計学に関する 2 章以降の議論で必要な内容を説明している。2 章以降では、2-4 章までは KDE の議論を行い、5-7 章まではノンパラメトリック回帰の 1 種である局所多項式回帰 (LPR: local polynomial regression) の議論を展開している。8 章は本稿の結論にあたり、本稿の研究成果と今後の展望・課題についてまとめている。

## 方向統計学におけるカーネル密度推定量 (1.2 節, 2 章, 3 章, 4 章)

### 通常のカーネル密度推定量 (1.2 節)

実数値データのための標準的な KDE が満たす理論的性質を 3 つ簡単に述べる。

- (1) カーネル  $K$  が  $p$  次オーダーカーネルであるならば、平均積分二乗誤差 (MISE: mean integrated squared error) が 0 に収束する収束レートは  $\text{MISE} = O(n^{-2p/(2p+1)})$  である。
- (2) 2 次オーダーカーネルから高次オーダー KDE を構成可能である。
- (3) 平滑化パラメータ推定量の収束レートを導出可能である。

(1) の性質で述べた MISE とは、 $f$  の間の誤差を測る誤差基準であり、MISE の値が小さいほど KDE の推定精度は高いことを意味する。また、 $p$  次オーダーカーネルとは、カーネル  $K$  のモーメント  $\alpha_j(K) := \int x^j K(x) dx$  に関して  $0 < j < p$  次の  $\alpha_j(K)$  が 0 となるようなカーネルのことを言う。つまり、(1) の性質は、高次の  $p$  次オーダーカーネルを用いることで、MISE の収束レートを  $\text{MISE} = O(n^{-2p/(2p+1)})$  に改良可能なことを意味する。

(2) の性質の高次オーダー KDE とは、 $\text{MISE} = O(n^{-2p/(2p+1)})$  を達成可能な KDE のことを意味する。通常の KDE は、Jones and Foster (1993) による加法型構成法や Terrell and Scott (1980) による乗法型構成法を用いれば、2 次オーダーカーネル (または KDE) から、高次オーダー KDE を構築できる。これが (2) の性質の内容である。

(3) の性質の意味は、KDE の最適な平滑化パラメータの推定量の収束レートが導出され、推定

量の一致性が保証されていることがある。収束レートが導出されている代表的な推定量として最小二乗クロスバリデーション法 (LSCV: least squares cross validation) 法やダイレクト・プラグイン法 (DPI: direct plug-in rule) が挙げられる。

## Di Marzio 型カーネル密度推定量 (2 章)

2 章では Di Marzio et al.(2011) が提案した Di Marzio 型 KDE の理論的性質に関する議論を行った。Di Marzio 型 KDE とは、Di Marzio 型カーネル族を採用した KDE のことを言う。その特徴はカーネルのモーメントを sine 型モーメント  $\eta_j(K_\kappa)$  とし、低次の sine 型モーメントが 0 となる sine 型  $p$  次オーダーカーネルを用いた点にある。

2 章の中心的成果は、代表的な円周上の密度関数の一つである巻き込みコーシー (WC) カーネルを採用した KDE の漸近正規性とその MISE を導出したことである。Sine 2 次オーダーカーネルである巻き込みコーシー (: wrapped Cauchy) カーネルは、その MISE の収束レートが  $O(n^{-2/3})$  となる。これは 0 次オーダー KDE の収束レートに過ぎない。同じ sine 2 次オーダーカーネルに含まれるフォン・ミーゼス (VM: von Mises) カーネルの MISE の収束レートは  $O(n^{-4/5})$  であり、2 次オーダー KDE の収束レートである。両者の収束レートの違いは、同じ sine 2 次オーダーカーネルであってもその MISE の収束レートは異なることを示している。Di Marzio et al.(2011) 自身も高次の sine 型  $p$  次オーダーカーネルの中にバイアスを改良できないものが含まれていることを指摘している。この 2 つのことから (1) の性質が成り立たないことが分かる。

また、Di Marzio et al.(2011) は、VM カーネルに関しては 4 次以上の高次オーダー KDE を構築可能であることを示している。しかし、ほかのカーネルに関しても (2) の性質が成り立つかは分かっていない。方向統計学の KDE において平滑化パラメータに関する理論的性質を議論した研究は、筆者の知る限り存在しておらず、Di Marzio 型 KDE は (3) の性質を満たさない。

## Hall 型カーネル密度推定量 (3,4 章)

3 章と 4 章の目的は、(1)–(3) の性質を満たす方向統計学における KDE を考案することである。筆者は、Di Marzio 型 KDE の問題点は Di Marzio 型カーネルと sine  $p$  次オーダーカーネルを採用したことにあると考えて、Hall 型カーネル族を採用した Hall 型 KDE を提案した。

3 章は Hall 型 KDE の MISE の改良方法について議論した。Hall 型カーネル族の良い性質は、KDE のバイアスを高次の項まで展開したとき、その各項が平滑化パラメータ  $\kappa$  の級数  $\kappa^l$  と Hall 型カーネル族のモーメント  $\mu_l(L) := \int_0^\infty L(r)r^{(l-1)/2}dr$  の積から構成される点である。本稿は、 $0 < l < p$  次のモーメント  $\mu_l(L)$  が 0 となるような新しい  $p$  次オーダーカーネルを定義することで、 $p$  次オーダーカーネルに関するバイアスを改良可能なことを示した。また、この性質

を用いて、 $p$  次オーダーカーネルを採用したときの MISE が  $\text{MISE} = O(n^{-2p/(2p+1)})$  となることを証明した。また、Hall 型 KDE は、加法型構成法と乗法型構成法を 2 次オーダーカーネル (KDE) に適用することで高次オーダー KDE を構築できる。以上の議論から Hall 型 KDE は、(1) と (2) の性質を満たすことが分かる。

4 章では Hall 型 KDE に関する平滑化パラメータ推定量の理論的性質を議論した。4 章で扱った推定量は LSCV 推定量と DPI 推定量である。LSCV 推定量の収束レートは  $O(n^{-1/10})$  となり、DPI 推定量の収束レートは  $O(n^{-4/15})$  となることを明らかにした。3 章と 4 章の議論から Hall 型 KDE は (1)–(3) の性質をすべて満たすことが示された。

## 方向統計学におけるノンパラメトリック回帰 (1.3 節, 4 章, 5 章, 6 章)

### 通常のノンパラメトリック回帰 (1.3 節)

実数値データのための LPR に関する理論的性質について簡単に 3 つ述べる。

- (4) LPR の多項式の次数  $p$  が 1 であるならば、 $\text{MISE} = O_p(n^{-4/5})$  となる。
- (5) 高次の次数  $p$  を選択すれば MISE を改良可能である。
- (6) MISE が次元数に依存するという「次元の呪い」の性質を持つ。

(4) と (5) は LPR の説明変数が 1 次元のときの性質であり、(6) は LPR の説明変数が多次元のときの性質である。標準的な LRR は誤差基準として重み付条件付 MISE を用いているが、重み付条件付 MISE のことを単に MISE と表すことにする。

(4) の性質は、1 次のときの LPR を局所線形回帰 (LLR: local linear regression) と呼ぶが、LLR の MISE の収束レートは  $O_p(n^{-4/5})$  となり、2 次オーダー KDE の収束レートに対応する。(5) の性質の詳細を説明する。LPR の次数  $p$  が奇数のときは  $\text{MISE} = O_p(n^{-(2p+2)/(2p+3)})$  となり、次数  $p$  を上げることで MISE を改良可能である。詳細は省略するが、次数  $p$  が偶数のときも MISE を改良できる。(6) の性質は、次元の呪いと呼ばれる性質であり、標準的な LLR は説明変数が  $d$  次元ベクトルであるとき、その MISE の収束レートは  $O_p(n^{-4/(d+4)})$  となるので、説明変数の次元  $d$  が増えるごとに MISE の収束レートは遅くなり推定精度が悪化することを意味する。

### Di Marzio 型局所多項式回帰 (5 章, 7.1 節)

5 章では方向統計学における局所多項式回帰の一つである sine 型 LPR を扱った。主に、Di Marzio et al. (2009) が与えた Di Marzio 型 LPR の理論的性質について議論している。Di

Marzio 型 LPR は、Di Marzio 型カーネルを採用した sine 型 LPR のことを指す．特に，5 章は，説明変数がスカラーな角度変数であるときを対象とし，Di Marzio 型 LLR( $p = 1$  の場合) に VM カーネルと WC カーネルを適用したときの理論的性質を与えた．

Di Marzio 型 LPR は VM カーネルを適用した場合は  $\text{MISE} = O_p(n^{-4/5})$  となり，これは (4) の性質を満たす．一方，WC カーネルを適用したときの MISE の収束レートは  $O_p(n^{-2/3})$  となり，これは (4) の性質を満たさない．また，高次の場合に関する Di Marzio 型 LPR の MISE の収束レートは未解決問題であり，(5) の性質を満たさない．7.1 節で議論したように Di Marzio 型 LLR は，説明変数がトーラス  $\mathbb{T}^q$  上で与えられるベクトルや実数・円周  $\mathbb{R}^d \times \mathbb{T}$  上で与えられるとき，その MISE の収束レートは不明であり，(6) の性質を満たさない (トーラス  $\mathbb{T}^q$  と実数・円周  $\mathbb{R}^d \times \mathbb{T}$  の空間に関しては，図 1 を参照にすること)．

## Hall 型局所多項式回帰 (6 章)

6 章では，sine 型 LPR に非負の値を取る Hall 型カーネル族を適用した Hall 型 LPR の理論的性質を議論した．6 章では 5 章と同様に説明変数がスカラーな角度変数であるものを扱った．

本稿では，Hall 型カーネル族が実数直線上のカーネルに分布収束する性質を明らかにした．この性質を用いることで Hall 型 LPR の MISE を導出し，その収束レートを計算できる．実際に次数  $p$  が奇数のとき，MISE の収束レートは  $O_p(n^{-(2p+2)/(2p+3)})$  となる ( $p$  が偶数の場合は省略)．つまり， $p = 1$  のときは  $\text{MISE} = O_p(n^{-4/5})$  であるが，高次の  $p$  を選ぶことで MISE の収束レートを改善できる．したがって，Hall 型 LPR は (4) と (5) の性質を満たす．

## 方向統計学における多変量線形回帰 (7 章)

7 章では，方向統計学における多変量 LLR の理論的性質の解明に取り組んだ．対象としたのは，説明変数が実数・トーラス  $\mathbb{R}^d \times \mathbb{T}^q$  上の  $d + q$  次元ベクトルとなる実数・トーラス上の LLR である．実数・トーラス上の LLR は，Di Marzio et.al(2009) のトーラス  $\mathbb{T}^q$  上の LLR や Qin et al. (2011) の実数・円周  $\mathbb{R}^d \times \mathbb{T}$  上の LLR を含む一般的な定義であるという特徴がある (図 1 を参照)．実数・トーラス上の LLR 上に Hall 型カーネル族を適用した LLR の MISE の収束レートは  $O_p(n^{-4/(d+q+4)})$  となる．実数・トーラス上の LLR は一般的な定義であるので，この結果は，表 1 でまとめたように，トーラス上・実数空間上・実数円周上の LLR の MISE の収束レートを含んだ一般的な形である．つまり，方向統計学における多変量 LLR も次元の呪いの性質を持つ．したがって，Hall 型 LPR は (6) の性質も満たすことが分かる．

また，7.3 節で実数・トーラス上の LLR を金沢大学角間キャンパスの電力需要データに応用する実データ分析を行った．比較するために Johnson and Wehrly (1978) のパラメトリック回帰を採用した．実数・トーラス上の LLR は，パラメトリック回帰と比べて，データが持つ局所的

な傾向も表現できる柔軟なモデルを与えることができた．そのため，実数・トーラス上の LLR は，パラメトリック回帰よりも決定係数  $R^2$  の値が高い．この結果から，実数・トーラス上の LLR は，パラメトリック回帰よりも説明力が高い回帰モデルを構築可能なことが分かる．

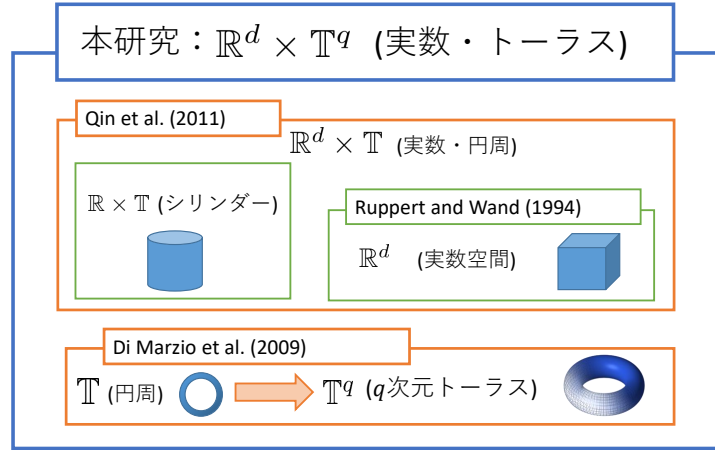


図 1: 実数・トーラス上の空間の概念図．提案する実数・トーラス上のノンパラメトリック回帰が先行研究を含むことを表している．

表 1: 多変量局所線形回帰 (LLR) の MISE の収束レート．

多変量 LLR の種類	標本空間	MISE の収束レート
実数・トーラス上の LLR	$\mathbb{R}^d \times \mathbb{T}^q$	$O_p(n^{-4/(d+q+4)})$
トーラス上の LLR	$\mathbb{T}^q$	$O_p(n^{-4/(q+4)})$
実数空間上の LLR	$\mathbb{R}^d$	$O_p(n^{-4/(d+4)})$
実数・円周上の LLR	$\mathbb{R}^d \times \mathbb{T}$	$O_p(n^{-4/(d+5)})$

## 結論 (8 章)

8 章では，7 章までの議論に基づいて，先行研究である KDE や LPR に Di Marzio 型カーネルを用いた Di Marzio 型 KDE・Di Marzio 型 LPR と比較することで，Hall 型カーネルを用いた Hall 型 KDE・Hall 型 LPR の優位性をまとめている (表 2-3 を参照)．つまり，本稿では，Hall

型カーネルを用いることで方向統計学におけるノンパラメトリック統計解析も実数値データのための通常のノンパラメトリック統計解析と同様な理論的性質を持つことを導出した。

また、実数・トーラス上の LLR を金沢大学角間キャンパスの電力需要データに応用する実データ分析を通して、方向統計学におけるノンパラメトリック統計解析は経済データの解析に有効であることを示した。

最後に今後の課題について述べる。本稿で提案した手法を周期的な変動を持つ経済データの解析に適用することで、方向統計学におけるノンパラメトリック統計解析手法がどのような経済現象の実証分析に貢献できるかを示すことが必要であろう。電力需要データなど周期的な変動を持つ経済データの多くは、同じ周期的変動を持つ時系列データと見なせる。したがって、計量経済学の視点から見たときの今後の研究課題は、方向統計学におけるノンパラメトリック統計解析手法を周期性を持つ時系列データの分析に応用できる可能性を、数理統計学的なアプローチで探ることである。

表 2: 方向統計学における KDE の性質の比較。○：当てはまる，×：当てはまらない，△：VM カーネルのみ当てはまる。

KDE	標準	Di Marzio 型	Hall 型
標本空間	$\mathbb{R}$ (実数)	$\mathbb{T}$ (円周)	$\mathbb{T}$
(1) : $p$ 次オーダーカーネル $\Rightarrow \text{MISE} = O(n^{-2p/(2p+1)})$	○	×	○
(2) : 2 次オーダーカーネルから高次オーダー KDE を構成	○	△	○
(3) : 平滑化パラメータ推定量の収束レートを導出可能	○	×	○

表 3: 標準的な LRR と Di Marzio 型 LPR の性質の比較。○：当てはまる，×：当てはまらない，△：VM カーネルのみ当てはまる。

LPR	標準	Di Marzio 型	Hall 型
(4): $p = 1 \Rightarrow \text{MISE} = O_p(n^{-4/5})$	○	△	○
(5): 高次の $p \Rightarrow \text{MISE}$ の収束レートの改善	○	×	○
(6): 次元の呪い (MISE が次元数に依存)	○	×	○

## 参考文献

- [1] Di Marzio, M., Panzera, A. and Taylor, C. C. (2009). Local polynomial regression for circular predictors. *Statistics & Probability Letters* **79**, 2066–2075.



- [2] Di Marzio, M., Panzera, A. and Taylor, C. C. (2011). Kernel density estimation on the torus. *Journal of Statistical Planning and Inference* **141**, 2156–2173.
- [3] Hall, P., Watson, G. S. and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, **74**, 751–762.
- [4] Jones, M. C. and Foster, P.J. (1993). Generalized jackknifing and higher order kernels. *Journal of Nonparametric Statistics* **3**, 81–94.
- [5] Johnson, R. A., and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, **73**, 602-606.
- [6] Qin, X., Zhang, J. S., and Yan, X. D. (2011). A nonparametric circularlinear multivariate regression model with a rule-of-thumb bandwidth selector. *Computers & Mathematics with Applications*, **62**, 3048-3055.
- [7] Terrell, G. R. and Scott, D. W. (1980). On improving convergence rates for nonnegative kernel Density Estimators. *The Annals of Statistics*, **8**, 1160-1163.

# 学位論文審査報告書

平成30年 1月23日

## 1 論文提出者

金沢大学大学院人間社会環境研究科

専攻 人間社会環境学

氏名 鶴田 靖人

## 2 学位論文題目（外国語の場合は、和訳を付記すること。）

方向統計学におけるノンパラメトリック統計解析

Nonparametric statistical analyses for directional statistics

## 3 審査結果

判定（いずれかに○印） ☒合格 ・ ☐不合格

授与学位（いずれかに○印） 博士（社会環境学・文学・法学・☒経済学・学術）

## 4 学位論文審査委員

委員長 寒河江 雅彦 ☒

委員 星野 伸明

委員 柳 在圭

委員 佐藤 清和

委員 澤田 幹

委員

（学位論文審査委員全員の審査により判定した。）

## 5 論文審査の結果の要旨

方向統計学とは角度データ(風向、24 時間周期で変動する電力需要量など)を扱う新しい統計学の分野である。角度データは単位円周上の確率分布に従う確率変数(角度変数)として扱われる。そのために、角度データを解析するには従来の統計手法をそのまま適用することはできず、方向統計学独自の統計理論が作られてきた。他方、複雑なデータ構造に柔軟に対応できるノンパラメトリック統計手法は、観測値ごとにカーネル関数(重み関数)を当てはめることで柔軟なモデリングが可能な特徴を持つ。方向統計学におけるノンパラメトリック統計解析の理論研究は未整備な状況にある。申請者は、方向統計学にノンパラメトリックな推定理論を構築するという高い目標の下に角度データに従う確率分布を推定するカーネル密度推定法と角度データを説明変数(例: 風向)とし実数変数を目的変数としたノンパラメトリック回帰モデルの推定理論においていくつかの研究成果を導いた。

具体的には、方向統計学におけるカーネル密度推定量の平均積分二乗誤差(MISE)を改良する理論の構築で Hall 型カーネル族を用いた新しいモーメントを定義し、推定精度を改善する高次オーダーカーネルを提案した。実証研究での利用が可能となる具体的な高次オーダーカーネルの構成法とカーネル関数に含まれる平滑化パラメータを推定する 2 つの方法に関して最小二乗クロスバリデーション法とダイレクト・プラグイン法による推定値の漸近正規性を示し、それぞれの平滑化パラメータの収束レートを求めた。

ノンパラメトリック回帰は Hall 型カーネル族を用いて、MISE の収束レートを導出した。特に説明変数がスカラーであるとき、ノンパラメトリック回帰の多項式の次数を上げることで MISE の収束レートが改善可能なことを証明した。また、説明変数が多次元角度変数ベクトルと多次元実数変数ベクトルからなる一般的な多次元ベクトルを考えて、この多変量ノンパラメ

トリック回帰の **MISE** の収束レートを与えた。申請者の提案した方向統計学におけるノンパラメトリック回帰モデルは、円周上、トーラス上、シリンダ上、球面上の角度情報を持つ幾何的な空間を統一的に扱うことのできる初めての回帰モデルである。

上記の論文は、2 編の査読付き論文誌(英分)に採択、2 編の査読付き論文誌（英文）（査読後の改訂中）で構成されている。ノンパラメトリック統計学の期待される若手研究者である。博士後期課程の3年間に学会発表、シンポジウム等での研究発表も10件を越え、活発な研究活動に邁進している。また、論文審査、審査委員との質疑でも的確な返答をしている。以上の研究業績、審査結果を踏まえ、博士（経済学）を授与することを認めます。